From Descriptive Analysis to Prescriptive Analytics: Visualization and Validation in Shared Print

Charleston Conference
November 8, 2017
Susan Stearns, Project Director
Mei Mendez, Project Manager
The EAST mission

Securing the scholarly print record in support of teaching, learning and research
- 60 members from Maine to Florida
- 52 Retention Partners
- 8 Supporting Partners
- Diversity in collection size, research vs. teaching focus and collection philosophy
Establish a safety net

- Secure scarcely held and frequently used
- Secure adequate copies to support demand
- Group agreement on retention models
Collection Analysis is Key

• For EAST: to identify titles to be committed to retention
• For individual libraries: to provide insight into de-selection and overall collection management
• Partnering with OCLC Sustainable Collection Services (SCS)®
• Analysis for original 40 Cohort 1 libraries completed in 2016
• Now undertaking analysis with 12 Cohort 2 libraries
• Scope of analysis is circulating print monographs only
A Comparison of the EAST Cohorts 1 and 2 Collective Collections

Over 2.2M holdings unique to EAST 2

- EAST 2
  - 12 libraries
  - 8.97M holdings
  - 62% held by 1 library

- EAST 1
  - 40 libraries
  - 16.5M holdings
  - 20% with >20 aggregate uses
  - 39% in HathiTrust
Uniqueness and overlap in the EAST collective collection – Cohort 1
Overlap by library – Cohort 2

- EAST 2 - Bucknell
- EAST 2 - Davidson
- EAST 2 - DeSales
- EAST 2 - Florida St.
- EAST 2 - Furman
- EAST 2 - Gettysburg
- EAST 2 - Hofstra
- EAST 2 - New York
- EAST 2 - Syracuse
- EAST 2 - U Pitt
- EAST 2 - Union Coll
- EAST 2 - U of South

Legend:
- Held by 1 group member
- Held by 2 group members
- Held by 3+ group members
Cohort 1 libraries only

**All Included Libraries**
- Retained Percentage: 36%
- Retained Holdings: 6,003,283

**Brandeis University**
- Retained Percentage: 36%
- Retained Titles: 286,045

**Criteria**

**LIBRARIES**
- Includes 36 of 36 Libraries

**RETAIN ALL**

**EAST holdings**
- FEWER THAN 5

**US holdings**
- FEWER THAN 40 (ANY EDITION)

**Regional Large Academic Libraries holdings**
- FEWER THAN 5 (ANY EDITION)

**SM Retention Rule**
- Retained Holdings
Validation Sample Study

What is the EAST Validation Sample Study?

A small sample study to determine an “availability metric” that provides a rough estimate of the percentage of library holdings that are likely to be available (are accounted for).

Why validate?

Trust.
What does the validation look like?

Worked with statistical consultant Grant Ritter and EAST Data Librarian Sara Amato to create an availability and condition matrix.

Determined that 6,000 items per library would be enough for a statistically valid sample.

SCS generated lists of 6,000 titles randomly selected from the in-scope records that were sent to them.
Validation Sample Study
Validation Sample Study Results

240,000
• Titles checked in the Validation Sample Study

92,575
• Of titles checked received retention commitments

70,000
• Of the full set of retained titles were potentially at risk of being in poor condition or not being on the shelf

Statistician reviewed titles that had a greater than 7.5% chance of being missing or in poor condition AND that EAST had surplus copies that were not already under retention.

Libraries with additional copies of these items not already under retention received requests to retain them.
Validation Sample Study Results: Missing Items

• Calculation of percentage of monographs missing for each participating EAST library

• Summary Statistics and Distribution on Missing Monographs at 40 participating EAST libraries (all estimates accurate to 1%)

<table>
<thead>
<tr>
<th>Rate</th>
<th>mean</th>
<th>std</th>
<th>Min</th>
<th>5th</th>
<th>25th</th>
<th>50th</th>
<th>75th</th>
<th>95th</th>
<th>max</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3.0%</td>
<td>2.1%</td>
<td>.3%</td>
<td>.6%</td>
<td>1.3%</td>
<td>2.8%</td>
<td>4.2%</td>
<td>7.4%</td>
<td>9.7%</td>
</tr>
</tbody>
</table>
Validation Sample Study Results: Condition

- Calculation of percentage of monographs in poor, acceptable, or excellent condition for each EAST library
- Summary Statistics and Distribution on Condition of Monographs at 40 participating EAST libraries (all estimates accurate to 2%):

<table>
<thead>
<tr>
<th></th>
<th>mean</th>
<th>std</th>
<th>Min</th>
<th>5th pctl</th>
<th>25th pctl</th>
<th>50th pctl</th>
<th>75th pctl</th>
<th>95th pctl</th>
<th>max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poor</td>
<td>10.7%</td>
<td>9.8%</td>
<td>1.1%</td>
<td>1.3%</td>
<td>3.0%</td>
<td>8.2%</td>
<td>15.2%</td>
<td>29.3%</td>
<td>44.2%</td>
</tr>
<tr>
<td>Acceptable</td>
<td>55.7%</td>
<td>17.3%</td>
<td>34.8%</td>
<td>34.8%</td>
<td>45.5%</td>
<td>54.1%</td>
<td>65.9%</td>
<td>88.1%</td>
<td>91.1%</td>
</tr>
<tr>
<td>Excellent</td>
<td>36.6%</td>
<td>17.7%</td>
<td>6.8%</td>
<td>7.8%</td>
<td>20.7%</td>
<td>33.6%</td>
<td>43.3%</td>
<td>70.5%</td>
<td>78.8%</td>
</tr>
</tbody>
</table>
Validation Sample Study Follow Up: Missing Items

• Missing rates at most participating libraries were very low. Only two libraries had rates greater than 7.4%, and no library had a rate greater than 10%

• Factors significantly affecting likelihood of being missing were small

• No factor appeared important enough to suggest a modification to the current EAST retention plan
• Poor condition rates at some participating libraries were large enough to merit attention. A quarter of the libraries had rates greater than 15% and two libraries had rates greater than 29%

• Certain factors affecting likelihood of being in poor condition are large enough to recommend changes to retention plans

• Retention plan might keep extra copies of older monographs, monographs frequently circulated, and monographs in subject area of painting
Validation Sample Study, Deeper Dive into Data

**Missingness:**
- If the missing volume is one of several retained copies, the problem is likely to be minimal. On average, the mean probability of two volumes of the same title being missing would be negligible.
- If the missing copy is the only copy of the monograph within EAST at allocation, then there is nothing to be done to mitigate the issue.
- However, 7,882 titles had unallocated copies at other EAST libraries. It is requested that an additional copy of these titles be retained.

**Condition:**
72,757 titles are more than 50% likely to be in poor condition and have unallocated copies at other EAST libraries. It is requested that an additional copy of these titles be retained.
Validation Sample Study, Cohort 2

- Professor Ritter determined that the same methodology and sample size could be applied to the Cohort 2 collections.
- EAST Cohort 2 libraries are conducting the validation sample study and expect to be completed in mid-December, when the data will be provided to Professor Ritter for analysis.
- The results of his analysis will be incorporated in the Cohort 2 retention modeling and used to select the final retention model.
Retention of Serials and Journals

• 29 of the 60 EAST libraries are Serials/Journals Retention Partners
• In late 2016 began to investigate options for collection analysis
• No decision support tools similar to GreenGlass® available today
• Contracted with the Center for Research Libraries for a series of reports on overlap and uniqueness as well as depth of holdings based on library contributed metadata
• No ability to compare outside of the EAST data set
Cohort 1 serials/journals titles as presented in Tableau for analysis: 108,863 title sets representing 195,793 holdings.
Totals, by library, of those titles held by 4-6 of the retention partners: 7,348 titles representing 34,405 holdings

<table>
<thead>
<tr>
<th>Library Title Sets</th>
<th>EAST Holdings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amherst</td>
<td>1,187</td>
</tr>
<tr>
<td>Bard</td>
<td>118</td>
</tr>
<tr>
<td>Boston College</td>
<td>4,535</td>
</tr>
<tr>
<td>Boston University</td>
<td>4,423</td>
</tr>
<tr>
<td>Brandeis</td>
<td>2,755</td>
</tr>
<tr>
<td>Connecticut College</td>
<td>1,592</td>
</tr>
<tr>
<td>Elms</td>
<td>44</td>
</tr>
<tr>
<td>Fairfield</td>
<td>372</td>
</tr>
<tr>
<td>Five Colleges</td>
<td>2,503</td>
</tr>
<tr>
<td>Hamilton</td>
<td>1,253</td>
</tr>
<tr>
<td>Lafayette</td>
<td>1,564</td>
</tr>
<tr>
<td>Loyola</td>
<td>598</td>
</tr>
<tr>
<td>Mount Holyoke</td>
<td>1,406</td>
</tr>
<tr>
<td>Siena</td>
<td>374</td>
</tr>
<tr>
<td>Smith</td>
<td>1,247</td>
</tr>
<tr>
<td>Swarthmore</td>
<td>832</td>
</tr>
<tr>
<td>Trinity</td>
<td>1,217</td>
</tr>
<tr>
<td>UMass Amherst</td>
<td>3,993</td>
</tr>
<tr>
<td>Union</td>
<td>1,219</td>
</tr>
<tr>
<td>Williams</td>
<td>2,406</td>
</tr>
<tr>
<td>Woods Hole</td>
<td>767</td>
</tr>
</tbody>
</table>

EAST Holdings (total holdings for these titles in EAST)
Library Title Sets (how many titles in my library)
Understanding your local collection – an EAST example

SCS has been delivering a set of Key Metrics to every client library since the launch of GreenGlass in 2012. For each measure, we offer a green hyperlink to the corresponding percentage of your overall collection. The accumulation of these measures provides an interesting opportunity for you to compare your own key measures with most of the others worked. These numbers are based on more than 220 US academic libraries and over 98 million holdings. They include research libraries, state universities, 4-year colleges, and 2-year colleges.

<table>
<thead>
<tr>
<th>METRIC</th>
<th>FILTERED ITEMS</th>
<th>% OF YOUR FILTERED ITEMS</th>
<th>SCS AVERAGE</th>
<th>COMPARED TO THE RANGES AND ADVICE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zero recorded uses</td>
<td>511,471</td>
<td>42% •</td>
<td>42% **</td>
<td></td>
</tr>
<tr>
<td>Greater than 3 recorded uses</td>
<td>305,404</td>
<td>25% •</td>
<td>25% **</td>
<td></td>
</tr>
<tr>
<td>Publications more than 10 years old</td>
<td>1,049,533</td>
<td>86% •</td>
<td>88% **</td>
<td></td>
</tr>
<tr>
<td>More than 100 US holdings - same edition</td>
<td>677,230</td>
<td>56% •</td>
<td>74% **</td>
<td></td>
</tr>
<tr>
<td>Fewer than five US holdings - any edition</td>
<td>18,031</td>
<td>1% •</td>
<td>2% **</td>
<td></td>
</tr>
<tr>
<td>Unique in New York - any edition</td>
<td>52,782</td>
<td>4%</td>
<td>NA</td>
<td></td>
</tr>
</tbody>
</table>
Titles published before 1900 with minimal usage by LC class
Comparing the local collection to the consortial collective data set
Descriptive Analysis and Visualizations

Descriptive analyses provide measures of composition and distribution of data.
Prescriptive Analytics

Prescriptive analytics assist you in finding the best course of action in a given situation.
Thank you.

sstearns@blc.org

mmendez@blc.org